

UNE MÉTHODE DE DÉTECTION MULTIÉCHELLE FONDÉE SUR LE MAXIMUM DE VRAISEMBLANCE ET LES ONDELETTES MULTIPLES : DÉFINITION ET PERFORMANCES

Zyed TIRA^(1,2), Patrick DUVAUT⁽¹⁾, Agnès JOUSSELLIN⁽²⁾

(1) ETIS-ENSEA 6 avenue du Ponceau 95014 Cergy-Pontoise cedex

(2) EDF/DER/SDM/TAI 6 quai Watier 78401 Chatou cedex

RÉSUMÉ

Nous présentons dans ce papier un algorithme de détection-segmentation des variations élémentaires dans la moyenne d'un signal. Cet algorithme détermine la décomposition optimale du signal observé sur des ondelettes multiples. Pour ce faire, on utilise l'approche du maximum de vraisemblance généralisé appliqué aux différentes échelles de la décomposition. On fusionne ensuite la décision en analysant la cohérence multiéchelle. Cet algorithme est illustré sur un signal test. Ses performances sont analysées à l'appui des courbes COR.

ABSTRACT

We present in this paper an algorithm of detection and segmentation of elementary variations in the average of a signal. This algorithm determines the optimal decomposition of the observed signal on multiple wavelets. To make this, one uses the approach of the generalized maximum likelihood applied on the different scales of the decomposition. One merges then the decision by analyzing the multiscale coherence. This algorithm is illustrated on a test-signal. Its performances are analyzed using the ROC curves.

1. INTRODUCTION

On donne dans ce document la définition, les performances ainsi qu'une illustration d'un algorithme de détection-segmentation multiéchelle de transitoires. Cet algorithme utilise le principe du maximum de vraisemblance généralisé [WIL76] pour déterminer la décomposition optimale du signal observé sur des ondelettes multiples. Il s'agit de déterminer à chaque échelle (décomposition pseudo-continue) une partition optimale de la fenêtre d'observation dont le nombre de segments est inconnu. Pour ce faire, on commence par décomposer le signal observé sur quatre ondelettes de référence ; le résultat de la projection est appelé fonction d'ambiguïté. Les ondelettes de référence sont choisies de manière à approcher au mieux les transitoires élémentaires qui nous intéressent. Le seuil de la fonction d'ambiguïté et des amplitudes estimées ainsi que la fusion de l'information multiéchelle permet de détecter et segmenter les différents transitoires. Cette fusion effectue une analyse de la cohérence des maximums locaux. Ce détecteur fournit en outre une estimation de l'amplitude des transitoires potentiels.

On se propose de détecter et de segmenter des variations élémentaires dans la moyenne d'un signal [BAS93] (cf. figure 1). Le terme élémentaire désigne ici la monotonie de la variation. On suppose que ces variations sont modélisables par un ensemble de variations de référence. Le terme modélisable est à prendre au sens large, c'est-à-dire qu'une variation élémentaire pourra être reconstituée par un élément de l'ensemble de référence, à une erreur quadratique près. On suppose en outre que les paramètres d'échelle, d'amplitude, de ligne de base et de temps

d'arrivée sont inconnus et que le signal est noyé dans un bruit blanc gaussien.

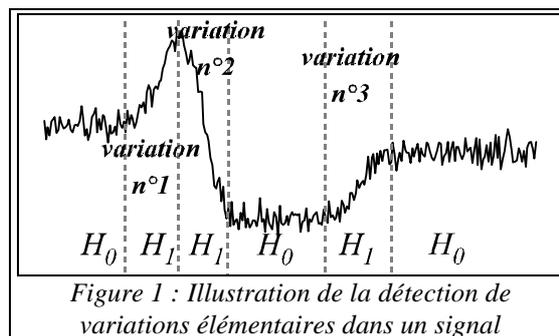


Figure 1 : Illustration de la détection de variations élémentaires dans un signal

On désigne par segmentation le fait de :

- déterminer les dates de début et de fin d'une variation monotone noyée dans du bruit,
- séparer deux variations monotones consécutives et de signes opposés.

Le problème ainsi défini est un problème de détection-estimation conjointe multi-hypothèses. Dans le cas où le bruit seul est présent, c'est l'hypothèse H_0 qui est décidée. Si une variation élémentaire est présente, c'est l'hypothèse $H_{1,i}$ correspondant à l'élément de l'ensemble de référence le plus vraisemblable qui est retenue. Dans la mesure où seule l'information concernant la présence ou non d'une variation nous intéresse, nous conservons pour $i > 0$, l'hypothèse $H_{1,i}$ la plus vraisemblable que l'on notera H_1 .

On utilise une stratégie du type estimation préalable qui consiste à tester simultanément les m hypothèses $H_{1,1}, \dots, H_{1,m}$. À chaque instant t , on conserve l'hypothèse qui est localement la plus vraisemblable en échelle. $H_{1,j}$ étant l'hypothèse choisie, on teste alors H_0 contre $H_{1,j}$ (détection de présence). Les transitoires recherchés doivent avoir une

amplitude significative (supérieure à un seuil). On confirme alors la décision par rapport à l'amplitude estimée au sens du maximum de vraisemblance (MV).

Les observations étant entachées de paramètres inconnus, nous utilisons la fonction de vraisemblance généralisée (FVG) pour réaliser la décision. Le processus de décision applique un seuil sur les fonctions de vraisemblance généralisées d'une part, et sur l'amplitude de la variation éventuelle estimée au sens du MV d'autre part. Pour fusionner les résultats des seuils, on utilise une stratégie "et" : une variation est classée comme variation utile si sa FVG et son amplitude estimée au sens du MV dépassent les seuils minimums respectifs. Ces seuils dépendent de l'échelle et sont fixés en fonction de l'a priori dont on dispose.

Dans les signaux qui nous intéressent, le bruit est plutôt présent sur les grandes échelles (premières voies) tandis que les variations (ou transitoires) utiles sont plutôt présentes sur les échelles intermédiaires et les petites échelles. Il s'avère par conséquent naturel d'imposer un seuil sévère pour la FVG sur les grandes échelles et d'assouplir ce seuil quand l'échelle décroît. D'autre part, le fait d'adopter un seuil en amplitude dépendant de l'échelle nous permet de détecter des variations dont la dynamique minimale est différente en fonction de l'échelle optimale (voie sur laquelle la variation est la mieux visible).

2. FORMALISATION DU PROBLÈME

L'ensemble E des formes de référence est constitué par les quatre ondelettes de la figure 2.

On s'intéresse aux transitoires résultant d'une concaténation ou d'une superposition d'un ou de plusieurs éléments de E .

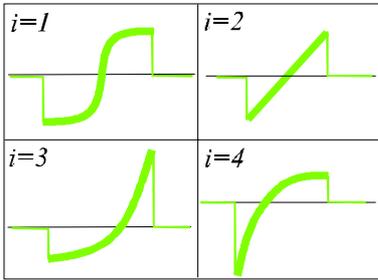


Figure 2 : Ensemble E des variations de référence

Notons $(u[n])_{n \in \mathcal{S}}$ le signal observé, où $\mathcal{S} = \{1, 2, \dots, N\}$. On suppose que les signaux $s_i(t)$ de l'ensemble de référence E sont centrés. On cherche à déterminer la partition optimale de \mathcal{S} , notée $(\mathcal{S}_k)_{1 \leq k \leq K}$, telle que

pour $n \in \mathcal{S}_k, \mathcal{S}_k = \{t_{0_k}, \dots, t_{0_k} + \beta_k T - 1\}$,

$$\begin{cases} u_c^{(k)}[n] = w[n] & \text{si } H_0 \\ u_c^{(k)}[n] = \alpha_k \cdot s_{i_k} \left[\frac{n - t_{0_k}}{\beta_k} \right] + w[n] & \text{si } H_{1,i_k} \end{cases} \quad (6)$$

- α_k et β_k sont les facteurs d'amplitude et d'échelle,
- w est un bruit blanc gaussien centré, de variance σ^2 ,
- $u_c^{(k)}$ est la version centrée de la restriction de u à \mathcal{S}_k :

$$\text{pour } n \in \mathcal{S}_k, u_c^{(k)}[n] \stackrel{\Delta}{=} u[n] - \left(\frac{1}{\beta_k T} \sum_{l \in \mathcal{S}_k} u[l] \right) \quad (7)$$

- T est la taille des variations de référence à l'échelle 1.

3. ALGORITHME

On se propose dans ce qui suit de détailler les différentes étapes de la mise en œuvre de l'algorithme de détection-segmentation multiéchelle.

1) On désire déterminer toutes les occurrences d'éléments de l'ensemble E dans le signal observé u . On désignera par composante de E dans u et on notera $C_{uE}(t_0, \beta)$ une telle

occurrence ayant lieu à l'instant t_0 et à l'échelle β . Les composantes les plus vraisemblables sont lues sur les maximums locaux de la fonction de décision $\hat{\chi}(t_0, \beta)$ où

$$\hat{\chi}(t_0, \beta) \stackrel{\Delta}{=} \chi(\hat{i}_{MV}(t_0, \beta), t_0, \beta) \quad (8)$$

$$\hat{i}_{MV}(t_0, \beta) \stackrel{\Delta}{=} \arg \left\{ \max_i [\chi(i, t_0, \beta)] \right\}$$

$$\text{et } \chi(i, \beta, t_0) = \frac{\int u_{c,0}^*(t) \cdot s_i \left(\frac{t - t_0}{\beta} \right) dt}{\left[\int \left| s_i \left(\frac{t - t_0}{\beta} \right) \right|^2 dt \right]^{1/2} \cdot \left[\int |u_{c,0}(t)|^2 dt \right]^{1/2}}$$

2) Les variations ainsi extraites peuvent correspondre à une bonne détection (hypothèse H_1) ou à une fausse alarme (H_0). Afin de trancher entre les deux hypothèses, nous comparons $\hat{\chi}(t_0, \beta)$ ainsi que le résultat de l'estimation de l'amplitude à des seuils dépendant de l'échelle β :

$$\left| \hat{\chi}(t_0, \beta) \right| \underset{H_0}{>} \eta(\beta) \quad \text{et} \quad \left| \hat{\xi}(t_0, \beta) \right| \underset{H_0}{>} \lambda(\beta) \quad (9)$$

$$\text{où } \hat{\xi}(t_0, \beta) = \frac{\int u_{c,0}(t) \cdot s_{\hat{i}_{MV}(t_0, \beta)} \left(\frac{t - t_0}{\beta} \right) dt}{\int \left| s_{\hat{i}_{MV}(t_0, \beta)} \left(\frac{t - t_0}{\beta} \right) \right|^2 dt} \times \text{amp} \left(s_{\hat{i}_{MV}(t_0, \beta)} \right)$$

et $\text{amp} \left(s_{\hat{i}_{MV}(t_0, \beta)} \right)$ désigne l'amplitude (= maximum - minimum) du signal de référence $s_{\hat{i}_{MV}(t_0, \beta)}$.

Remarque : Dans la suite on suppose que les signaux sont réels et on omet donc la marque de conjugaison *.

La fonction $\hat{\chi}(t_0, \beta)$ mesure la vraisemblance de l'hypothèse H_1 relativement à l'hypothèse H_0 . Elle prend ses valeurs dans l'intervalle $[-1, 1]$. Son signe fournit le sens de la variation en question :

$$\hat{\chi}(t_0, \beta) > 0 \Leftrightarrow \text{variation croissante,}$$

$$\hat{\chi}(t_0, \beta) < 0 \Leftrightarrow \text{variation décroissante.}$$

Notons toutefois que la propriété de croissance ou décroissance a un sens uniquement quand $\hat{\chi}(t_0, \beta)$ est assez grand en valeur absolue. Dans le cas contraire, il s'agit plutôt de bruit dont la moyenne est à tendance croissante ou décroissante.

Une première manière de choisir les seuils $\eta(\beta)$ et $\lambda(\beta)$ consiste à fixer la probabilité de fausse alarme associée à chacun des tests de l'équation (6) et à déterminer les seuils qui maximisent les probabilités de détection respectives. Cette opération nécessite la connaissance de l'écart-type du bruit, ce qui n'entre pas dans nos hypothèses.

Nous adoptons par conséquent une démarche empirique pour fixer ces seuils. On remarque que lorsque la quantité $(1-\eta)$ décroît, la probabilité de fausse alarme décroît. D'autre part, pour une probabilité de détection donnée, la probabilité de fausse alarme décroît en fonction de la voie d'analyse. On choisit donc de prendre un seuil $\eta(\beta)$ qui décroît en fonction de la voie et qui soit proche de 1 sur les premières voies. L'assouplissement du seuil $\eta(\beta)$ quand l'échelle décroît permet de tenir compte du fait que la modélisation des variations du signal observé par des éléments de E n'est pas exacte.

Le seuil en amplitude est possible et bénéfique lorsqu'on possède un a priori sur la valeur minimale de l'amplitude du signal. On prend alors le seuil λ égal à cette amplitude minimale. Quand on est en présence d'un signal dont l'amplitude minimale des variations élémentaires dépend de la voie de meilleure observabilité, on prend un seuil $\lambda(\beta)$ qui dépend de l'échelle.

C'est par exemple le cas lorsque le signal observé est constitué d'une concaténation aléatoire de motifs tels que ceux donnés dans la figure 3. On se place dans le cas où le motif M_1 est rapide (mieux visible sur les premières voies) alors que le motif M_2 est lent.

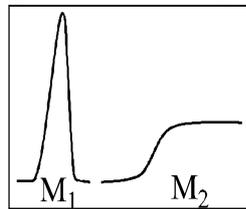


Figure 3 : Exemple de motifs élémentaires

On suppose que les amplitudes minimales respectives A_1 et A_2 des motifs M_1 et M_2 vérifient $A_1 \gg A_2$ (par exemple $A_1 = 2A_2$). L'allure du seuil λ en fonction du paramètre d'échelle est alors celui de la figure 4. Ceci permet de diminuer la probabilité de fausse alarme sur les premières voies sans dégrader la probabilité de détection.

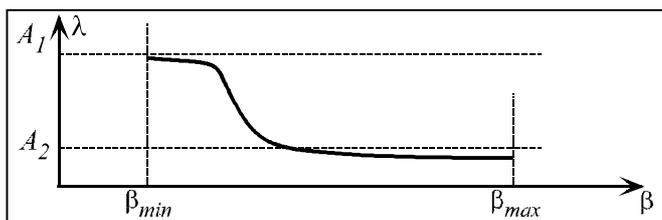


Figure 4 : Exemple de définition d'un seuil en amplitude

3) Soit $\Lambda_s(\beta) = \left\{ \left(t_{0,s}^k(\beta), \beta, \hat{\chi}(t_{0,s}^k(\beta), \beta) \right), 1 \leq k \leq K_s(\beta) \right\}$ la suite des maximums locaux en valeur absolue (MVA) de

$\hat{\chi}(t_0, \beta)$ correspondant à l'hypothèse H_1 (ayant donc subi avec succès l'opération précédente de seuillage). Chacun de ces maximums correspond à une variation élémentaire $C_{u|E}(t_{0,s}^k(\beta), \beta)$ de date de début $t_{0,s}^k(\beta)$, de durée $\beta \times T$ et d'amplitude estimée $\hat{\xi}(t_{0,s}^k(\beta), \beta)$. On appelle recouvrement les situations dans lesquelles deux variations successives sur une même voie empiètent l'une sur l'autre, c'est-à-dire

$$\exists k \in \{1, \dots, K_s(\beta) - 1\} \text{ tel que, } t_{0,s}^k(\beta) + \beta \times T \geq t_{0,s}^{k+1}(\beta) \quad (10)$$

On élimine cette ambiguïté en conservant à chaque occurrence de recouvrement la variation la plus vraisemblable (c'est-à-dire celle qui rend maximum $\hat{\chi}(t_0, \beta)$). Ceci revient à chercher, pour une échelle donnée, la variation la plus vraisemblable.

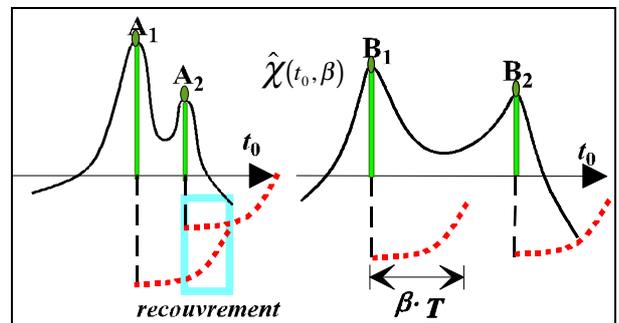


Figure 5 : Illustration de l'élimination des recouvrements

A_1, A_2, B_1 et B_2 sont des MVA de $\hat{\chi}(t_0, \beta)$ à l'échelle β . En pointillés sont tracés les transitoires (ici les ondelettes) correspondant. Dans le cas des MVA A_1 et A_2 , on observe un empiètement entre les transitoires correspondant. On élimine alors le MVA correspondant au transitoire le moins vraisemblable, soit ici A_2 . En revanche, dans le cas des MVA B_1 et B_2 , il n'y a pas de recouvrement : les deux MVA sont alors conservés.

4) Soit $\Lambda_r(\beta) = \left\{ \left(t_{0,r}^k(\beta), \beta, \hat{\chi}(t_{0,r}^k(\beta), \beta) \right), 1 \leq k \leq K_r(\beta) \right\}$

la suite des MVA de $\hat{\chi}(t_0, \beta)$ après élimination des recouvrements (traitement 3). On effectue alors une analyse de cohérence multiéchelle afin d'éliminer les variations parasites. Il s'agit des variations dont le support temporel est inclus dans celui d'une variation de signe opposé et qui se trouve sur une voie inférieure :

si $\exists \beta_1(\beta_2)$ et $\exists (k_1)k_2)$ tels que

$$\left\{ \begin{array}{l} \left[t_{0,r}^{k_1}(\beta_1), t_{0,r}^{k_1}(\beta_1) + \beta_1 \times T \right] \subset \left[t_{0,r}^{k_2}(\beta_2), t_{0,r}^{k_2}(\beta_2) + \beta_2 \times T \right] \\ \text{et} \\ \left(\hat{\chi}(t_{0,r}^{k_1}(\beta_1), \beta_1) \times \hat{\chi}(t_{0,r}^{k_2}(\beta_2) + \beta_2) \right) < 0 \end{array} \right. \quad (11)$$

alors on élimine l'extremum (et donc la variation)

$$\left(t_{0,r}^{k_1}(\beta_1), \beta_1, \hat{\chi}(t_{0,r}^{k_1}(\beta_1), \beta_1) \right).$$

Cette opération permet d'augmenter la robustesse au bruit mais, en contrepartie, elle peut causer la non-détection de variations telles que la variation A de la figure 6. Notons toutefois que cette opération n'est pas indispensable pour

l'algorithme de détection. Elle peut en effet être omise si l'on est amené à traiter des variations du type de la figure 6 au prix d'une détérioration de la robustesse au bruit.

5) Les traitements précédents nous ont permis d'identifier les composantes les plus vraisemblables du signal observé dans le plan temps-échelle. On se propose dans ce qui suit de fusionner l'information entre les différentes échelles. Cette fusion consiste à prendre, à un instant donné, le point de la variation la plus vraisemblable en échelle.

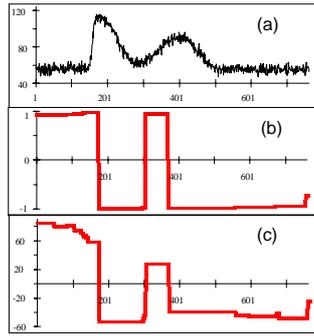


Figure 6 : Résultat du détecteur multiéchelle sur un signal-test

$$\text{Soit } \Lambda_f(\beta) = \left\{ \left(t_{0,f}^k(\beta), \beta, \hat{\chi} \left(t_{0,f}^k(\beta), \beta \right) \right), 1 \leq k \leq K_f(\beta) \right\}$$

les MVA à l'échelle β qui ont été retenus à l'étape précédente. Considérons la fonction à deux variables définie comme suit :

Pour $\beta_{\min} \leq \beta \leq \beta_{\max}$

Pour $1 \leq k \leq K_f(\beta)$

Pour $t_{0,f}^k(\beta) \leq t_0 < t_{0,f}^k(\beta) + \beta T$,

$$\Pi(t_0, \beta) = \hat{\chi} \left(t_{0,f}^k(\beta), \beta \right)$$

Si $\nexists k \in \{1, \dots, K_f(\beta)\}$ tel que $t_0 \in [t_{0,f}^k(\beta), t_{0,f}^k(\beta) + \beta T]$

alors $\Pi(t_0, \beta) = 0$

Le résultat de fusion multiéchelle est donné par la fonction

$$\text{suivante : } \hat{\chi}_f(t_0) = \max_{\beta} [\Pi(t_0, \beta)]$$

6) Les parties de signe constant de $\hat{\chi}_f(t_0)$ correspondent à des variations monotones (simples ou composées) du signal observé. Une variation est dite composée si elle est le résultat de la fusion de différentes variations entre les échelles ou de la concaténation d'éléments différents de l'ensemble E de même sens de variation, sur une échelle, où des deux à la fois. Par conséquent, la segmentation des parties de signes constants de $\hat{\chi}_f(t_0)$ nous permet de dénombrer les variations élémentaires présentes dans le signal observé.

Ces fonctions sont illustrées dans la figure 7 sur un signal-test. Leurs performances sont évaluées à l'aide de courbes COR pour trois valeurs de RSB (cf. figure 8).

4. APPLICATION DE L'ALGORITHME DE DETECTION

Dans (a) est donné le tracé d'un signal-test. Dans (b) est donnée la version de $\hat{\chi}(t_0, \beta)$ après seuillage et fusion multiéchelle noté $\hat{\chi}(t_0)$. Dans (c) est donné le signal de l'estimation de l'amplitude correspondant. La segmentation des signaux tracés dans (b) fournit le nombre de variations présentes dans le signal.

On obtient ainsi quatre variations. L'amplitude de chacune des variations détectées est lue sur les signaux c à l'abscisse qui rend maximum le segment correspondant dans le signal (b). Les résultats obtenus en ce qui concerne le nombre des variations et l'amplitude de chacune des variations sont conformes au nombre de variations et à l'amplitude effective du signal-test.

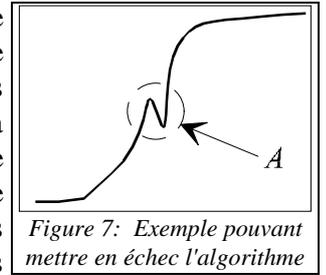


Figure 7 : Exemple pouvant mettre en échec l'algorithme

5. ANALYSE DES PERFORMANCES

Afin d'évaluer les performances, on calcule un nombre moyen de fausses alarmes pour un taux moyen de bonnes détections. Le graphique ainsi obtenu est appelé courbe COR modifiée. Le taux moyen de fausses alarmes est à valeurs dans l'intervalle [0,1] : il s'agit d'une estimation de la probabilité de fausse alarme. Pour calculer ce taux on procède comme suit : on fixe un taux moyen de bonnes détections ; ceci revient à fixer un seuil ; le taux de fausse alarme est égal au rapport entre les fausses alarmes qui dépassent ce seuil par le nombre total de fausses alarmes. On moyenne ensuite sur plusieurs réalisations du bruit pour obtenir le taux moyen. Le nombre moyen de fausses alarmes est à valeurs dans l'intervalle [0,N] où N est le nombre maximum de fausses alarmes observées pour une taille fixée du signal. Pour un taux moyen de bonnes détections on calcule le nombre total de fausses alarmes et on moyenne ensuite sur plusieurs réalisations du bruit pour obtenir le nombre moyen.

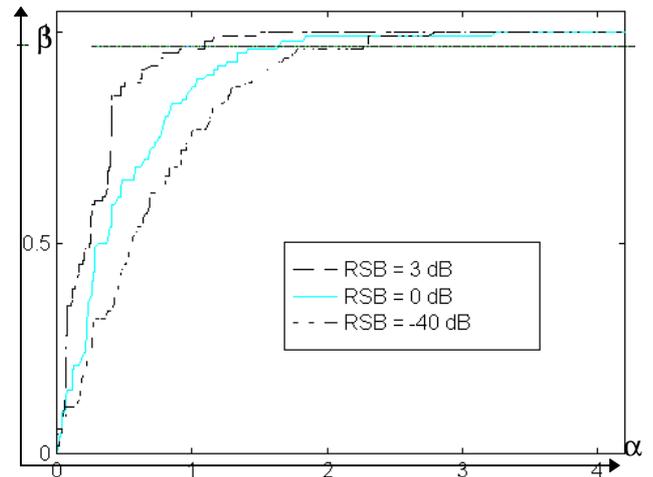


Figure 8 : Courbes COR modifiées α (resp. β) désigne le nombre moyen de fausse alarme (resp. taux moyen de détection)

RÉFÉRENCES

[BAS93] Michèle Basseville, Igor V. Nikiforov, "Detection of abrupt changes, theory and applications", Prentice Hall, Inc. 1993.
 [WIL76] A. S. Willsky and H. L. Jones, "A generalized likelihood ratio approach to the detection and estimation of jumps in linear systems," IEEE Trans. Automat. Contr., vol. AC-21, pp. 108-112, Feb. 1976.
 [TIR96] Zyed TIRA, "Apport de l'analyse multirésolution oblique et du rapport de vraisemblance généralisé à la reconnaissance partiellement coopérative", Thèse de doctorat, Université de Cergy, juin 1996.